DOCUMENT RESUME

ED 430 007 TM 029 720

AUTHOR Mott, Michael S.; Hare, R. Dwight
TITLE The Reliability and Developmental Validity of the Writing What You Read Rubric for Hypermedia-Authored Narratives: Grades 2-3.
PUB DATE 1999-04-00
NOTE 23p.; Paper presented at the Annual Meeting of the American Educational Research Association (Montreal, Quebec, Canada, April 19-23, 1999). For the assessment report, see TM 029 719.
PUB TYPE Reports - Research (143) -- Speeches/Meeting Papers (150)
EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Comparative Analysis; *Computer Assisted Instruction; Computer Software; Elementary Education; Elementary School Students; Evaluation Methods; *Hypermedia; Reliability; Test Construction; Test Format; *Test Use; Validity
IDENTIFIERS Paper and Pencil Tests; Scoring Rubrics; *Writing What You Read

ABSTRACT

This study investigated the reliability and developmental and concurrent validity of the Writing What You Read (WWYR) rubric, an instrument originally designed for use with paper-and-pen-created narratives, for hypermedia productions of students in grades 2 and 3. Four teachers guided their students in a 3-month-long hypermedia/process writing curriculum. The 60 students created narratives in HyperStudies 3.1 on Windows-based computers. In addition to test elements, the narratives contained audio, video, graphic, and hyper-travel elements. Five raters evaluated the hypermedia narratives using the WWYR. Results indicate that, for WWYR values as classroom tool, rater judgments were reliable and valid. However, it is noted that, for WWYR usefulness in guiding educational policy, the reliability and validity issues need to be further evaluated. (Contains 1 figure, 7 tables, and 29 references.) (Author/SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

Running head: RELIABILITY AND VALIDITY OF THE WWYR

ED 430 007

The Reliability and Developmental Validity of the Writing What You Read Rubric for Hypermedia-Authored Narratives: Grades 2-3

Michael S. Mott

National-Louis University

R. Dwight Hare

Mississippi State University

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
This document has been reproduced as received from the person or organization originating it.
Minor changes have been made to improve reproduction quality.
Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY
Michael Mott
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

Paper Presented at the 1999 Annual Meeting of the

American Educational Research Association

Montreal, Canada

April 19-23

TM029720

BEST COPY AVAILABLE

Abstract

This study[1] investigated the reliability and developmental and concurrent validity of the Writing What You Read (WWYR) Rubric, an instrument originally designed for use with paper and pen-created narratives, for hypermedia productions of students in grades 2 and 3. Teachers (η=4) guided their students in a three months-long hypermedia/process writing curriculum. Students (η=60) created narratives in HyperStudio 3.1 on Windows-based computers. In addition to text, the narratives contained audio, video, graphic and hyper-travel elements. Raters (η=5) evaluated the hypermedia narratives using the WWYR. Results indicated that, for WWYR value as a classroom tool, rater judgment scores were reliable and valid. However, it was noted that, for WWYR utility in guiding educational policy the reliability and validity issue needs to be further evaluated.

Background

Beginning in the 1960s, researchers created a theoretical model for the implementation of writing instruction based upon the experiential and socially driven learning theories of John Dewey (Langer & Allington, 1992). Dewey theorized that effective learning takes place when the learner draws from prior experience to acquire new knowledge.

Theoretical models of writing instruction first designed in the 1960s by Rohman and Emig (1971, as cited in Reed, 1996) were built upon two older models that stressed a process learning approach: (a) the Three Stage Model, consisting of pre-writing, drafting and revising; and (b) the Recursive Model, in which the non-linear nature of writing was revealed (Reed, 1996). Graves (1983) and Calkins (1983) further supported the view that elementary writing should be taught as a process and recommended Writing Workshop, a process of writing instruction based upon both the Three Stage and Recursive Models of writing instruction. Writing Workshop continues to receive research-based support and further definition as an effective model for process writing instruction (Atwell, 1987; Calkins & Harwayne, 1991; Graves & Sunstein, 1992; Sinicki, 1996; Strech, 1994).

The curriculum changes in writing instruction in the 1960s required alternative assessments to measure the whole essay (e.g. papers consisting of narrative, expository, persuasive, as well as other genres) and not just portions of an essay (Huot, 1990). Prior to the 1970s, most writing was assessed indirectly by scoring students' writing through the counting of isolated mechanical features (e.g. syntax, word counts, grammar checks, etc.), rather than directly (judgment scores on the overall merit, or quality, of an entire essay). In the early 1970s, however, the practice of directly assessing writing became widely adopted even though there were problems with this form of assessment. For example, Huot critiqued the attempt to have two or more raters reliably rate the quality of an essay

[1] The current research report is based on Mott (1998), an unpublished dissertation. Mott and Halpin (1999) provide a different view of the current study by placing results in a technical context.

and discussed the need for further research in this area even though raters achieved adequate levels of interrater reliability.

More recently, direct assessments evaluating the quality of student work have appeared in rubric form. The Writing What You Read (WWYR) rubric (Wolf & Gearhart, 1994) has been extensively evaluated and offered as a direct writing assessment for elementary writing instruction (see Table 1). This rubric has been found to be reliable and valid as a measure of the quality of student writing in paper and pen writing environments (Gearhart, Herman, Novak & Wolf, 1995; Novak, Herman & Gearhart, 1996). A comparable rubric for the measurement of the quality of students writing (e.g., coherence, cohesiveness and overall merit of the writing) in computer-based writing environments, however, has yet to be developed and evaluated (Kinzer & Leu, 1997).

Computers are transforming the way many of us read, write and think. As a result, the tools and definitions of literacy are undergoing extensive change (Costanzo, 1994). With the proliferation and increasing popularity of the microcomputer, writing is no longer confined to a boundary set by paper. On paper the writer cannot instantaneously edit, revise or control font size and type as one can on a microcomputer. Moreover, varied forms of multimedia and hypermedia programs have been introduced into the computer writing environment (Palumbo & Prater, 1993). According to Dauite and Morse (1994), the rationale for engaging elementary students in writing within a multimedia environment is the enhancement of children's expression and communication, not only through text but also through the use of a variety of meaning-based symbol systems. Hypermedia provides the benefits of writing in a multimedia environment in addition to the choice of creating non-linear connections (Palumbo & Prater, 1993). Because of the inclusiveness of multimedia features in all hypermedia programs, this study will only refer to the term hypermedia.

Numerous higher-level cognitive tasks have been identified for student writing in hypermedia environments: organizing, reasoning, selecting and connecting (Yang, 1996). Since the advent of hypermedia, there is a pressing need for researchers to address the effect of student writing within hypermedia environments (Ayersman, 1996; Reed, 1996). Little research, however, has addressed the assessment of hypermedia-created student outcomes or productions (Kinzer & Leu, 1997; Palumbo & Prater, 1993). There is a need to examine the reliability and validity of pen and paper rubrics for non- paper environments (Gearhart et al. 1995).

The WWYR Rubric

Writing What You Read (WWYR) is an analytic trait narrative writing assessment developed for elementary students grades K-6 (see Table 1). (This section, unless cited otherwise, is based entirely on Wolf & Gearhart,1994). WWYR is a rubric assessment. The benefits of rubric assessments are that they:

> enable raters to apply standard criteria in making judgments about the quality of students' work, thus reducing the subjectivity of the scoring process and helping to assure that students' scores are based on their performance and not the idiosyncrasies of the individual raters. (Abedi, 1997, p. 8)

WWYR is a unique rubric in that it was specifically designed for the assessment of the narrative genre. Wolf and Gearhart stated their rationale for constructing a narrative writing assessment instrument as follows: "We want educators to teach narrative, not as an ever-shifting set of lovely stories to be lauded, but as a foundation for analysis, reflection, and criticism which can, in turn, be used as a resource for children's original writing" (Wolf & Gearhart, p. 5). An important rationale, set forth by the authors, for WWYR use by teachers is to facilitate genre-specific comments toward students who are creating narratives and to avoid generic comments such as, "nice work" (Wolf & Gearhart, 1994, p. 5).

The words Writing What You Read originated from commonly articulated teachers' advice for narrative writing which was to encourage students to "write about what you read" (p. 5) or have recently read or experienced. The purpose for using WWYR can thus be connected to current literacy theoretical frameworks that call for the need for children to experience literature-through reading, writing and other media. The overall goal for the application of WWYR was described from the context of the classroom:

> In our work with teachers, we ask them to analyze literature in terms of the following narrative components: genre, theme, character, setting, plot, point of view, style and tone. Teachers' understanding of the components of narrative then become a springboard for integrating curricular possibilities, instructional techniques, and assessment tools. Our goal is to help teachers assess children's narrative writing in the same way that they critically respond to literature. Equipped with the "tools of the literary trade"-an understanding of genre influences, the technical vocabulary, and the orchestration of the narrative components within a text-teachers can reflect on and offer their students explicit guidance for their writing. (p. 3)

Two tools were designed to support these goals: the Narrative Feedback Form and the WWYR Rubric.

WWYR Assessment Tools

The Narrative Feedback form (see Figure 1) was also designed Wolf and Gearhart (1994) to facilitate directed interaction between students and teachers "to provide space for constructive and critical comments in the narrative areas of Theme, Character, Setting, Plot and Communication, as well as in two issues generic to all writing-Convention and Writing Process" (p. 5). The form was also designed to be used as a record keeping device to document students' development in writing over time.

The WWYR Rubric contains five evaluative scales designed to assess students' developing competencies in narrative writing: Theme, Character, Setting, Plot and Communication. The vertical analytical evaluative scales (1-6 for each competency) were designed to enable teachers to make instructional decisions on specific narrative components (Theme, Character, Setting, Plot or Communication) a student needs reinforcement in, and were not intended as a method for assigning a numerical value to a narrative. (The numerical value has been used in educational research solely to examine the reliability and evidence for the validity of the WWYR). Teachers merely shade off a box in the rubric to denote where a child's narrative is along each competency. The authors stressed the importance of the analytic scales by stating that:

> an analytical response to narrative elements can be helpful for indicating both strengths and areas needing refinement. If a child writes a piece of realistic fiction with little or no conflict, underdeveloped characters, and no theme, there is certainly room for explicit guidance to lead the child toward more effective writing, and a helpful "next step" might indeed be a focus on enhancing the narrative's treatment of just one of the elements. Even mature, successful writers could look to the rubric to assess their stories and see possibilities for further development. (p. 15)

In addition to the vertical evaluative scales the authors have incorporated dual dimensions to accompany the scales to address the multitude of purposes found in the narrative genre. The dimensions were classified in the following manner:

> Themes move between explicit and sometimes didactic statements to implicit revelations. Characters can be flat personalities who remain static and unchanging in a story, or they can come equipped with more rounded physical and emotional description and change over time. The Setting can be a simple cardboard backdrop, or it can take on a more essential, multifunctional role. The Plot can also be simple and without tension, or it can evolve in conflict and complexity. Narrative Communication can move between literal and symbolic meanings in style and tone (p. 7)

The dual dimensions enable the WWYR Rubric to serve as an assessment tool for a wide array of narrative subgenres. The authors summarized the purpose of the dual dimensions in stating that "the dual dimensions are not only assessment tools but touchstones for selecting materials and designing lessons" (p. 8).

WWYR Overview

According to the authors the WWYR was designed to strengthen the connection between curriculum, instruction and assessment. The main purpose of this assessment is for teachers to explicitly guide their students in growing and maturing as narrative writers through reflection and analysis. Thus assessment of writing should focus and guide teachers to develop lessons that address writing areas in need of improvement. By leading students through an analysis of their writing, similar to the manner in which they analyze literature, their narrative writing abilities can be broadened.

Design

This study used an ex post facto design (Gay, 1996) with a comparative component (Gall, Borg & Gall, 1996) to examine the technical qualities of the WWYR for scoring hypermedia-created narratives. First, the degree of correlation among rater judgments along the five WWYR analytic sub-scales for student hypermedia-created narrative productions was determined. Second, the level of difference (evaluated using Hotelling's T^2) between WWYR subscale scores across grade levels (2 and 3) was determined. Third, the level of difference (evaluated using a One-Way MANOVA) between WWYR subscale scores was determined for low, medium and high ability groups, within the two third grade groups, where ability groups were operationally defined using ITBS (Linn & Willson, 1990) scores.

This design was based, in part, upon the model writing assessment study conducted by Gearhart et al. (1995). There are two differences, however. One, the Gearhart et al. study did not provide a time frame for their WWYR rater training procedures, whereas a time frame is provided in the current study. Two, the Gearhart et al. study compared two assessments. They compared one rubric, not yet scrutinized by research, with one research-tested rubric. This was done in order to provide evidence for the developmental validity of the WWYR rubric. In the current study existing literacy achievement scores (ITBS), designed to measure general literacy cognitive skills, were used to compare and evaluate the concurrent validity (i.e., the degree to which a test is related to an already established test) of the WWYR results. Thus, developmental and concurrent validity were addressed, not only through grade level comparisons, as they were in the Gearhart et al. study, but through comparisons with the reliable and valid ITBS assessment.

Similar designs have been used to examine the usefulness and technical qualities of the WWYR for pen and paper-created narratives (Gearhart et al. 1995) as well as for collections (portfolios) of student narratives (Novak et

al. 1996). This study had similar design advantages to those found in the studies by Gearhart et al.(1995) and Novak et. al (1996) described in Chapter Two. These designs used a three-step approach to examine the technical merits of the WWYR rubric: (a) percent agreement between raters, (b) paired correlations to determine interrater reliability, and (c) grade level comparisons to determine developmental and concurrent validity. In order to examine the technical qualities of the WWYR for hypermedia-created narratives, two questions were addressed. One, can raters reliably rate student narratives created in hypermedia using a rubric originally designed for pen and paper-created narratives? Secondly, will the data secured from the raters' support the validity (e.g., developmental and concurrent validity) of WWYR scores with regard to judging students' hypermedia-writing performance?

The implementation of this design enabled the researchers to build evidence for the developmental and concurrent validity of WWYR scores for judging student hypermedia-writing performance. The issue of validity was approached based upon Messick's (1992) interpretation of the hierarchical nature of validity. Developmental validity was defined by Messick (1992) as evidence for the score sensitivity of an assessment to the development of the learner in the specific domain being assessed. For example, if an assessment reflects lower scores for younger children and higher scores for older children, then that assessment would be considered sensitive to the chronological development of the students. In turn, to follow the assumption set forth by Messick, that assessment would also contain score sensitivity to the cognitive development of the younger and older students since older students often have greater cognitive skills than younger students. Gearhart et al. (1995) argued, citing Messick's definition of developmental validity, that for students in higher grades, writing assessment scores should be higher than for students in lower grades if the assessment is developmentally valid, or sensitive to the competency levels of the authors of the writing samples.

The design of this study also enabled the researchers to examine the concurrent validity of WWYR scores for judging students' hypermedia-created narrative productions. In his hierarchy, Messick placed concurrent validity below content-related validity. Thus, in order for a measure to achieve content-related validity it must first demonstrate acceptable concurrent validity. This feature of the design was important since the WWYR was an unproven measure for evaluating the literacy-based performance of the students' hypermedia-created productions.

Procedures

Teacher-participants, although familiar with the HyperStudio software environment and process writing curriculum, received 3 hours of training in HyperStudio programming prior to engaging their students in the

HyperStudio/process writing curriculum (for a description of HyperStudio software see Appendix A). (All procedures for students and teachers implemented in this study were previously created and evaluated in an unpublished pilot study conducted by the researcher in an 8 week summer program in 1997). HyperStudio training for the teachers consisted of completing all exercises in HyperStudio Express (Cochran & Staats, 1997). In addition to the HyperStudio exercises, teachers experienced a 1 hour HyperStudio/process writing curriculum training session that functioned to reacquaint them with the curriculum. This training session took place through an interactive hypermedia learning module "Teach Creative Writing with HyperStudio" (unpublished computer module by Mott, 1998). This module was used and evaluated in a hypermedia for instruction graduate level course taught at Mississippi State University on-line (http://www.coolteaching.com/HS/ syllabus.html).

Students experienced and experimented within the HyperStudio software environment a total of 6 times, in 45 minute sessions in the computer laboratory, prior to engaging in the HyperStudio/process writing curriculum. During these student training sessions, three exercises were completed: (a) Tool Use, (b) Writing with Text, and (c) Button Making. The students completed these three exercises with the remainder of the time spent in an open-ended forum practicing the experimentation and application of skills that were reviewed in the exercises.

Students' work in HyperStudio required the manipulation of a wide array of tools. The Browse Tool was used to navigate through the narrative productions by pointing and clicking. The Paint Brush enabled students to apply brush strokes, in 32 colors, to their backgrounds. The Eraser, Line Tool and Pencil were used to accent scanned-in drawings students completed on paper prior to engaging in HyperStudio programming. Lastly, the Paint Bucket enabled students to fill in spaces in their documents in solid colors.

Students' manipulation of Text Tools in HyperStudio centered on the Text Box. The Text Box is a box within a HyperStudio Card (page) that allows for the incorporation of text. Text Boxes function similarly to word processors. Erasing, cutting and pasting were accomplished as they can be in a basic word processor. Font size, style and color were manipulated by students to enhance their hypermedia narrative products. (Teachers required students to use darker colors for text to avoid difficulty in reading). In addition to Text Boxes, Graphics Text was used by students mainly for titles. Unlike Text Boxes, Graphics Text could not be erased, cut or pasted. Teachers insisted that students use Text Boxes for all text, except the title. (This allowed for revision and editing of all text, graphics and sound).

Hypermedia features (traveling, or linking, from "page to page" in a hypermedia document) were enacted by students' use of Buttons. Buttons were programmed to enable students to create "pages" they could navigate to in their hypermedia narrative products. In addition to the creation of "pages", students added sounds (bells, whistles, beeps, voice recordings, and music) to their button functions. A few students incorporated video images into their buttons. In their classrooms students experienced the entire five stages of process writing curriculum (brainstorming, drafting, revising, editing and publishing) in six 45 minute sessions, on paper, prior to composing in HyperStudio. Within a 3 week time-span, teachers and the researcher led students through nine 45 minute sessions in which students created HyperStudio narratives. The researcher was available during all sessions to address problems and concerns of teachers and students as they experience the curriculum. Teachers, with help from the researcher, assisted students with software problems (text, sound, and/or graphics). Students were not be evaluated by the researchers.

Raters

Several studies (Gearhart et. al, 1995; Novak et. al, 1996; Wolf & Gearhart, 1993a, 1994) conducted on the WWYR , for paper and pen-created narratives, have revealed that the training of raters can be accomplished in a brief time and during a combination training-rating. Unlike the Gearhart et al., Novak et al., and Wolf studies where all participants were experienced WWYR raters, in this study not all raters had the same level of expertise with WWYR scoring. Therefore, a 1 hour training session for inexperienced WWYR raters preceded a 2 hour rating and training session. The 1 hour training session took place in lecture and discussion format with all raters present.

One Hour WWYR Rater Training Session

The researcher used this hour to familiarize raters with the rationale for the assessment, defined each sub-scale (Theme, Character, Setting, Plot and Communication), and described the general characteristics of the HyperStudio narratives they rated using a lecture and discussion format. For the first 20 minutes, using a transparency of the WWYR rubric projected onto a projection screen, the researcher discussed each of the six value levels for all five subscales. During the middle 20 minutes of the first hour, raters were encouraged to identify concerns and ask questions regarding the subscales and the six levels in each subscale. The last 20 minutes of th training session were used for rater assessment of sample HyperStudio narratives. The HyperStudio narratives were projected onto the projection screen using an LCD panel and overhead projector to collaboratively establish benchmark scores along as wide a range of subscale values as possible. Raters scored "practice" samples (samples

not included in the data set) and then discussed the scores. Once the level of disagreement for all five raters was less than one point (± 1) along all five subscales for three HyperStudio narratives in succession scoring began. The training and rating sessions took place in the Instructional Resources Computer laboratory at Mississippi State University.

Rating Session Procedures

Each rater scored all randomly ordered samples of HyperStudio narrative productions. Raters were matched with a microcomputer. The following procedures were followed:

(1) Raters received all samples in the form of labeled diskettes. Grade levels of HyperStudio narrative productions were not revealed in order to avoid an interaction effect between scoring and grade level.
(2) Each sample had a two-digit number identifying the file which enabled the raters to open that file. The randomly ordered two-digit numbers also enabled the researcher to match the samples with the name, age, ITBS mean literacy score and classroom teacher of the author. Only the researcher knew which number corresponded to a particular child.
(3) Each rater was be provided with a stack of labeled WWYR rating sheets that contained their name and the randomly ordered sample to be rated on each sheet.
(4) Each rater viewed their own randomly ordered samples and thus viewed different HyperStudio narratives. A sample was viewed by choosing "Open" under the "File" command and selecting the appropriate sample number as labeled on the WWYR scoring sheet.
(5) Communication between raters took place during check-set breaks to allow for re-establishment of rating guidelines.
(6) Once a rater scored a sample, they placed their labeled WWYR scoring sheet underneath their stack and proceeded to the next sample.

Once all HyperStudio narrative samples were rated the session was complete.

Participants

Students

The student-participants in this study were second and third grade children, from four intact classrooms, from one elementary school in West Point, Mississippi. West Point is situated in the Northeast-Central part of the state. The student body was predominantly African-American and the students selected for the study represented the norm of the student body of the school. Students from two 3rd grade classes and two 2nd second grade classes were selected based upon their teachers' decision to implement the HyperStudio/process writing curriculum.

Teachers

Four teachers were selected based upon their agreement to teach the HyperStudio/process writing curriculum (see Table 2 for a conceptual summary of this curriculum). The four teachers participated in a summer technology and arts program and received HyperStudio and process-oriented writing curriculum training from the

researcher and another doctoral student involved in the program. For the 8 week program, each Friday was utilized as a teacher training and planning day, thus the participating teachers received several hours of training each week.

Raters

Five doctoral students were selected based upon their experience in elementary education, as well as in the teaching and assessment of student writing. The raters had either several years of teaching experience at the elementary level or several years of experience teaching language arts curriculum. Additionally, four of the raters had several years of experience teaching HyperStudio software and process-oriented writing curriculum to pre and inservice teachers in courses, workshops and consultations.

Results

WWYR Hypermedia Reliability

The percentages of agreement for the WWYR Rubric subscale scores averaged across ten pairs of raters were higher than those observed in two other WWYR studies (Gearhart et. al, 1995 and Novak et. al, 1996). See Table 3 for a summary of these results. Additionally, the Pearson correlations for the WWYR Rubric subscale scores averaged across ten pairs of raters revealed acceptable levels of interrater reliability for four of the five WWYR subscales: Theme, Character, Plot and Communication. Interrater reliability for the Setting subscale did not fall within the acceptable range of values. See Table 4 for a summary of these results. All WWYR subscale ratings were highly correlated, in both this study and in the Gearhart et. al study, which indicated that raters were not making highly differentiated judgments regarding the hypermedia narrative productions along the different subscales or dimensions addressed in the rubric. That is, the raters appear to basically evaluate the overall "quality" of the productions.

WWYR Hypermedia Developmental Validity

Raters judgment scores did not significantly differentiate students from grade 2 and grade 3 (Hotelling's T^2) $F(1,54)+.87$, $p.=.16$. Table 5 summarizes these results. Classroom variables and the fact that only students in only two grades (i.e., grades 2 and 3) participated in the current study may have contributed to the results, which revealed no significant differences in the scores assigned to the productions of students in grades 2 and 3. However, it is important to note that, although insignificant, students in grade 3 received higher ratings than students in grade 2 across all five WWYR subscale scores.

WWYR Hypermedia Concurrent Validity

The significant differences revealed between low, medium and high ITBS groups and WWYR subscale scores provided evidence for the sensitivity of the WWYR to the development of students' hypermedia/writing competence. Tables 6 and 7 summarize these results. The significant results of the One-Way MANOVA provided evidence that raters' judgments were evaluating students' skills as message-producers (communication through text and other meaning-based symbol systems). Dauite and Morse (1994), who used a similar curriculum (hypermedia/writing) in their study, found that students who were given the opportunity to compose in hypermedia were engaged in problem solving as they expressed themselves through the manipulation of a variety of meaning-based symbol systems, including text. Dauite and Morse drew the conclusion that students' hypermedia productions represented significant problem solving efforts, similar to what is required in process writing environments. The One-Way MANOVA did not yield results that would enable the researcher to directly describe the degree of relatedness of raters' WWYR judgments and students' ITBS scores. In order to describe the relationship between WWYR scores and literacy skill (as measured by the ITBS), an additional analysis was conducted.

Conclusion

Since the goals of classroom assessment for use by teachers can be vastly different than those for educational policy makers, it is important to qualify the reliability and validity results of the WWYR for use with hypermedia productions. The value of an assessment (the degree to which an assessment enhances teaching and learning in the classroom) dictates whether that assessment improves learning. The utility of an assessment (the degree to which a writing assessment reliably and validly can be applied to large-scale applications) dictates whether that assessment can inform educational policy makers (Wolf & Gearhart, 1994). Since assessment should inform educational policy, it is vital that a measure meet the necessary technical standards so that results can reliably and validly inform large-scale decisions (educational expenditures, etc.). On the other hand, it is vitally important that the same assessment have value for classroom use to improve teaching and learning.

There are two implications for the large-scale application of WWYR results for hypermedia. First, the low reliabilities revealed in this study, although acceptable for classroom use, may not be appropriate for large-scale assessment. Additionally, the unacceptable reliability found for the Setting subscale matched the Gearhart et al. (1995) finding and provides further evidence that the WWYR, for both pen and paper and hypermedia, needs to be improved in order to be a reliable large-scale measure.

Second, the content-related validity of the WWYR for hypermedia was not completely established. Messick (1992) argued that several validity types must be evaluated in order for a measure to have content-related validity. Developmental and concurrent validity represented two lesser validity types in his hierarchy. In order for the WWYR to be used for large-scale assessment of students' hypermedia products, other types of validity should be documented as well.

Recommendations

The results of this study suggest the following topics for additional research:

1. Further research of established writing rubrics and newly designed rubrics that measure hypermedia narrative productions are necessary to better understand the influence of hypermedia features on rater judgment.

2. A repetition of this study with an increased number of raters (generalizability study) should be undertaken to determine whether reliability will be increased for judgments of hypermedia narrative product quality.

3. A repetition of this study involving wider range of age groups (e.g., grades 1-6) should be undertaken to determine the reliability and developmental validity of WWYR rater judgments of hypermedia narrative products.

4. A repetition of this study using a research design that would allow the researcher to control student use of hypermedia features should be undertaken in order to evaluate how different hypermedia features (audio, video and graphics) influence raters.

5. A study of the effect of hypermedia programming on students' writing achievement should be undertaken.

6. A study of the effects of hypermedia learning on the concurrent learning in other content areas (e.g., science, math, and social studies) should be undertaken.

References

Atwell, N. (1987). In the middle. Portsmouth, N.H: Heinemann.

Calkins, L.C. (1983). Lessons from a child. Melbourne: Heinemann Educational Books.

Calkins, L.C. (1986). The art of teaching writing. Portsmouth, N.H: Heinemann.

Calkins, L.C., & Harwayne, S. (1991). Living between the lines. Portsmouth, N.H: Heinnemann.

Cochran, D.W., & Staats, R.A. (1997). HyperStudio express. New York: International Thompson Publishing Co.

Computer software review: HyperStudio. (1994, July). School Library Journal, pp. 53-55.

Costanzo, W. (1994). Reading, writing and thinking in an age of electronic literacy. In C.L. Selfe & S. Hilligoss (Eds.). Literacy and computers (pp. 11-21). New York: The Modern Language Association of America.

Daiute, C., & Morse, F. (1994). Access to knowledge and expression: Multimedia writing tools for students with diverse needs and strengths. Journal of Special Education Technology, 8(3), 221-256.

Gall, D.G., Borg, W.R., & Gall, J.P. (1996). Educational research: An introduction. London: Longman.

Gay, L.R. (1996). Educational Research. (5th ed.). New Jersey: Prentice-Hall.

Gearhart, M., Herman, J.L., Baker, E.L., & Whittaker, A.K. (1992). Writing portfolios at the elementary level: A study of methods for writing assessment (CSE Tech. Rep. No. 337). Los Angeles: University of California, Center for Research on Evaluation, Standards and Student Testing.

Gearhart, M., Herman, J.L., Novak, J.R., & Wolf, S.A. (1995). Toward the instructional utility of large-scale writing assessment: Validation of a new narrative rubric. Assessing Writing, 2(2), 207-242.

Graves, D.H. (1983). Writing: Teachers and children at work. Portsmouth, NH: Heinemann Educational Books.

Graves, D.H., & Sunstein, B.S. (Eds.). (1992). Portfolio portraits. Portsmouth, NH: Heinemann Educational Books.

Huot, B. (1990). The literature of direct writing assessment: Major concerns and prevailing trends. Review of Educational Research, 60(2), 237-263.

Kinzer, C., & Leu, D.J. (1997). Focus on research: Exploring literacy and learning in electronic environments. Language Arts, 74, 126-135.

Langer, S.A., & Allington, R.L. (1992). The history of writing and reading curriculum. In P.W. Jackson (Ed.). Handbook of research on curriculum (pp. 325-332). New York: Macmillan.

Linn, R., & Willson, V. (1990). Review of the Iowa Test of Basic Skills form J. In J. Conoley and J. Impara (Eds.), Mental Measurement Yearbook (9th ed.). University of Nebraska Press.

Messick, S. (1992). Validity of test interpretation and use. In M. Alkin (Ed.), Encyclopedia of education research (6th ed., pp. 1487-1495). New York: Macmillan.

Mott, M. (1998). Teach Creative Writing with HyperStudio for the Macintosh personal computer [Unpublished computer software]. Starkville, Mississippi.

Novak, J.R., Herman, J.L., & Gearhart, M. (1996). Establishing validity for perfomanced-based assessments: An illustration for collections of student writing. Journal of Educational Research, 89(4), 220-233.

Palumbo, D.P., & Prater, D. ((1993). The role of hypermedia in synthesis writing. Computers and Composition, 10(2), 59-70.

Reed, W.M. (1996). Assessing the impact of computer-based writing instruction. Journal of Research on Computing in Education, 28(4), 418-437.

Siniki, C.A. (1996). A study of the relationship of writing instruction, staff development practices, and writing performance of selected third grade classrooms (Doctoral dissertation, Northern Illinois University, 1996). Dissertation Abstracts International, 5709-A, 3820.

Strech, L.L. (1994). The implementation of writer's workshop: A review of the literature. Long Beach, CA: California State University. (ERIC Document Reproduction Service No. ED 380797).

The CRESST Line. (1997). Interrater reliability for performanced-based assessment. [Brochure]. Los Angeles, CA: Abedi, J.

Wagner, R. (1995). HyperStudio Workshop 3.0. El Cajon, CA: Roger Wagner Publishing.

Wolf, S.A., & Gearhart, M. (1993a, 1993b). Writing What You Read: Assessment as a learning event (CSE Tech. Rep. No. 358). Los Angeles: University of California, Center for Research on Evaluation, Standards, and Student Testing.

Wolf, S.A., & Gearhart, M. (1994). Writing What You Read: A framework for narrative assessment. Language Arts, 7(6), 425-445.

Yang, S.C. (1996). A dynamic reading-linking-to-writing model for problem solving within a constructive hypermedia learning environment. Journal of Educational Multimedia and Hypermedia, 5(3/4), 263-283.

Table 1

Writing What You Read Rubric

Narrative Rubric

Theme

explicit ◄——► implicit
didactic ◄——► revealing

- Not present or not developed through other narrative elements
- Meaning centered in a series of list-like statements ("I like my mom. And I like my dad. And I like my....") or in the coherence of the action itself ("He blew up the plane. Pow!")
- Beginning statement of theme—often explicit and didactic ("The mean witch chased the children and she shouldn't have done that."); occasionally the theme, though well stated, does not fit the story
- Beginning revelation of theme on both explicit and implicit levels through the more subtle things characters say and do ("He put his arm around the dog and held him close. 'You're my best pal,' he whispered.")
- Beginning use of secondary themes, often tied to overarching theme, but sometimes tangential; main theme increasingly revealed through discovery rather than delivery, though explicit thematic statements still predominate
- Overarching theme multi-layered and complex; secondary themes integrally related to primary theme or themes; both explicit and implicit revelations of theme work in harmony ("You can't do that to my sister!", Lou cried, moving to shield Tasha with her body.)

Character

flat ◄——► round
static ◄——► dynamic

- One or two flat, static characters with little relationship between characters; either objective (action speaks for itself) or first person (author as "I") point of view
- Some rounding, usually in physical description; relationship between characters is action-driven; objective point of view is common
- Continued rounding in physical description, particularly stereotypical features ("wart on the end of her nose"); beginning rounding in feeling, often through straightforward vocabulary ("She was sad, glad, mad.")
- Beginning insights into the motivation and intention that drives the feeling and the action of main characters often through limited omniscient point of view; beginning dynamic features (of change and growth)
- Further rounding (in feeling and motivation); dynamic features appear in the central characters and in the relationships between characters; move to omniscient point of view (getting into the minds of characters)
- Round, dynamic major characters through rich description of affect, intention, and motivation; growth occurs as a result of complex interactions between characters; most characters contribute to the development of the narrative; purposeful choice of point of view

Setting

backdrop ◄——► essential
simple ◄——► multi-functional

- Backdrop setting with little or no indication of time and place ("There was a little girl. She liked candy.")
- Skeletal indication of time and place often held in past time ("once there was..."); little relationship to other narrative elements
- Beginning relationship between setting and other narrative elements (futuristic setting to accommodate aliens and spaceships); beginning symbolic functions of setting (often stereotypical images—forest as scary place)
- Setting becomes more essential to the development of the story in explicit ways: characters may remark on the setting or the time and place may be integral to the plot
- Setting may serve more than one function and the relationship between functions is more implicit and symbolic—for example, setting may be linked symbolically to character mood ("She hid in the grass, clutching the sharp, dry spikes, waiting.")
- Setting fully integrated with the characters, action, and theme of the story; role of setting is multifunctional—setting mood, revealing character and conflict, serving as metaphor

Plot

simple ◄——► complex
static ◄——► conflict

- One or two events with little or no conflict ("Once there was a cat. The cat liked milk.")
- Beginning sequence of events, but occasional out-of-sync occurrences; events without problem, problem without resolution, or little emotional response
- Single, linear episode with clear beginning, middle, and end; the episode contains four critical elements of problem, emotional response, action, and outcome
- Plot increases in complexity with more than one episode; each episode contains problem, emotional response, action, outcome; beginning relationship between episodes
- Stronger relationship between episodes (with the resolution in one leading to a problem in the next); beginning manipulation of the sequence through foreshadowing, and subplots
- Overarching problem and resolution supported by multiple episodes; rich variety of techniques (building suspense, foreshadowing, flashbacks, denouement) to manipulate sequence

Communication

context-bound ◄——► reader-considerate
literal ◄——► symbolic

- Writing bound to context ("You have to be there) and often dependent on drawing and talk to clarify the meaning; minimal style and tone
- Beginning awareness of reader considerations; straightforward style and tone focused on getting the information out; first attempts at dialogue begin
- Writer begins to make use of explanations and transitions ("because" and "so"); literal style centers on description ("sunny day"); tone explicit
- Increased information and explanation for the reader (linking ideas as well as episodes); words more carefully selected to suit the narrative's purpose (particularly through increased use of detail in imagery)
- Some experimentation with symbolism (particularly figurative language) which shows reader considerations on both explicit and implicit levels; style shows increasing variety (alliteration, word play, rhythm, etc.) and tone is more implicit
- Careful crafting of choices in story structure as well as vocabulary demonstrate considerate orchestration of all the available resources; judicious experimentation with variety of stylistic forms which are often symbolic in nature and illuminate the other narrative elements

Walt Carlson/Quality Writing 1995

BEST COPY AVAILABLE

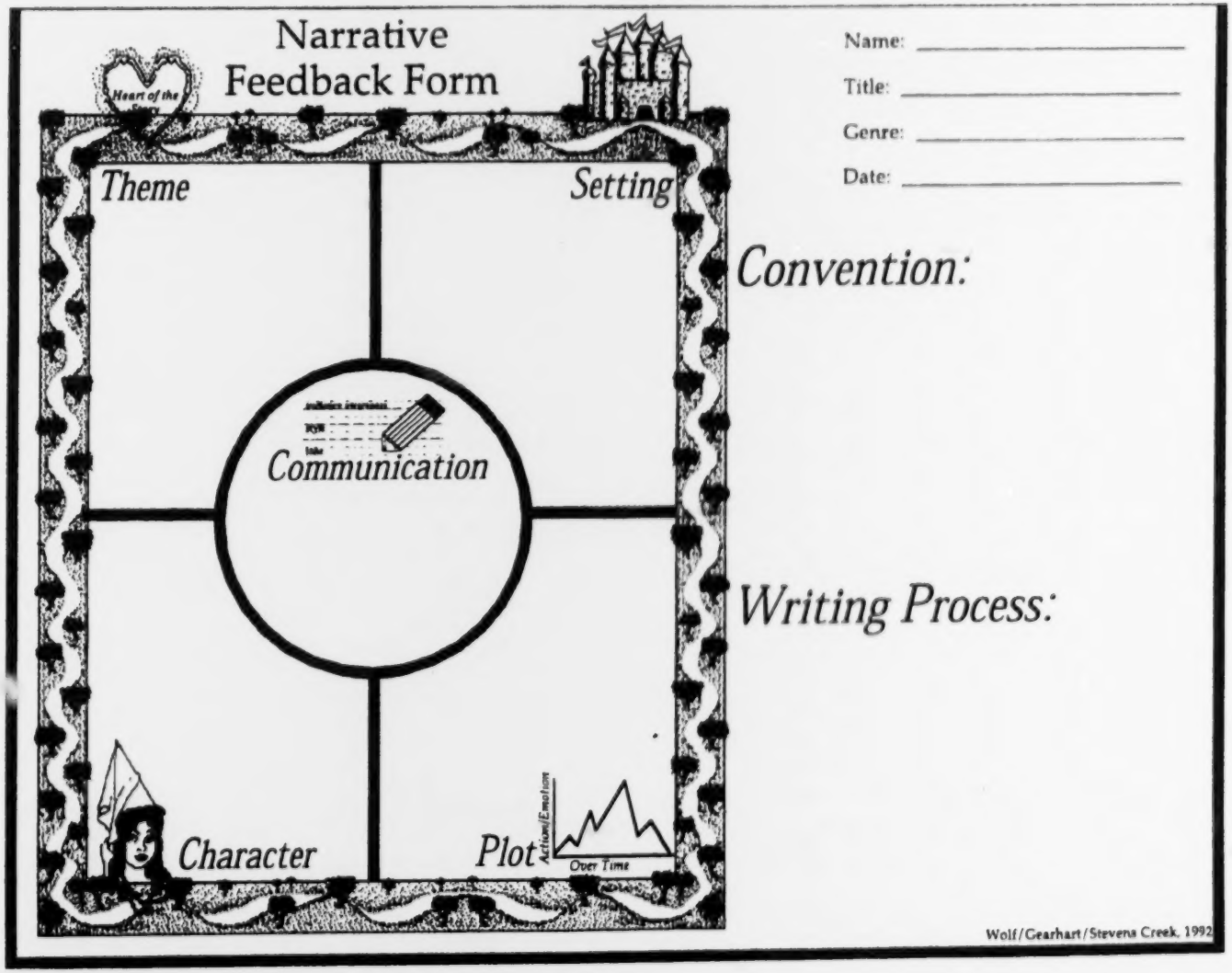

Figure 1

Narrative Feedback Form

Table 2

Hypermedia/Process Writing Curriculum Framework

Brainstorming	**Rough Draft**	**Revising**	**Publishing/Presenting**
Lead a discussion on possible topics that models brainstorming. Students then engage in brainstorming, followed by the choosing of one story topic. Students must then complete a storyboard (blank squares representing HS cards) that outlines the structure of their story.	Students create a draft of their story in HyperStudio. The teacher and/or a peer conference to review the draft and provide critical feedback. On a "Feedback Form" the HS-author notes what changes will be implemented. Draft two is then completed.	Students present the second draft two their teachers and the teacher assist in revising the HS-story for, not only writing mechanics (grammar, punctuation, spelling, etc.), but for other hypermedia features as well: audio, video, graphics, hyper connections, etc.	Students present their published hypermedia stories to the whole group in an "Authors Chair." Questions and answers follow. For presenting, the use of a large monitor or television connected to a computer may enhance the presentations.

Appendix A

HyperStudio Description

According to Cochran and Staats (1997) Hyperstudio software represents a hypermedia/multimedia authoring system for all age-groups that supports the inclusion of several elements in addition to text: graphics, sound, video and CD-ROM in a variety of ways. The Computer Software Review (1994) further described the multimedia features and related how they can be utilized:

> Users create individual computer screens called "cards." The program provides tools for adding graphics, sound, video and text to the cards. "Painting includes the usual Macintosh tools-pencil, paintbrush, spray paint, line, fill, rectangle, oval and polygon. Graphics can be imported from other sources or brought in with the numerous databases included with the program (p. 54).

In the Computer Software Review (1994) it was also noted that, for the feature of text, there are five font types and four font sizes available in addition to basic editing and revising features found in a word processor.

Multimedia features can be used by authors to create non-linear or linear presentations with links to "cards." Links and other multimedia features can be applied through "buttons", which can be created in a simple manner (Beekan & Beekman, 1993):

> Button-creation dialogue boxes offer more than color; they automate many processes that require scripting in HyperCard. A few mouse-clicks create a button that can play an automated sequence or QuickTime movie, activate a self-scrolling text field, or control a laser disc or CD (p. 81).

Table 3

Percentages of Agreement for all Five Subscales of the WWYR Rubric Averaged Across Ten Pairs of Raters

WWYR Subscale	*±0*	*±1*	*n*
Theme	.70	.96	60
Character	.78	.99	60
Plot	.73	.99	60
Setting	.67	.99	60
Communication	.68	.99	60

Table 4

Percentages of Agreement for the WWYR Rubric Averaged Across All Subscales

WWYR Rating Material and Grade	*±0*	*±1*	*n*
Hypermedia Narratives: Grades 2-3 Mott, 1998	.71	.98	60
Pen and Paper Narratives: Grades 1-6 Gearhart et al. 1995	.46	.96	120
Collections of Pen and Paper Narratives: Grades 2-5 Novak et al. 1996	.25	.94	52

Table 5

Descriptive Statistics: WWYR Subscales Across Grade Level

Statistics	*Dependent Variables*					
	n	*Theme*	*Character*	*Setting*	*Plot*	*Comm.*
Mean Vectors						
Grade Level						
2	20	2.25	1.90	2.14	2.19	2.18
3	40	2.62	2.21	2.30	2.47	2.46
Variance-Covariance Matrix						
Theme		.25	.28	.19	.21	.23
Character		--	.31	.20	.22	.21
Setting		--	--	.25	.22	.17
Plot		--	--	--	.29	.23
Comm.		--	--	--	--	.24

Table 6

Descriptive Statistics: WWYR Subscales Across Grade Level

Statistics	*Dependent Variables*					
	n	*Theme*	*Character*	*Setting*	*Plot*	*Comm.*
Mean Vectors						
Grade Level						
2	20	2.25	1.90	2.14	2.19	2.18
3	40	2.62	2.21	2.30	2.47	2.46
Variance-Covariance Matrix						
Theme		.25	.28	.19	.21	.23
Character		--	.31	.20	.22	.21
Setting		--	--	.25	.22	.17
Plot		--	--	--	.29	.23
Comm.		--	--	--	--	.24

Table 7

Mean WWYR Subscale Scores for Low, Medium and High

Ability Grade 3 Students

WWYR Subscale	*ITBS NPR/Literacy Category*	*Mean Score*	*SD*	*n*	*F*	*Sig*
Theme	Low	2.31	.62	16		
	Medium	2.80	.28	10	6.19	.01
	High	2.86	.31	13		
Character	Low	1.80	.50	16		
	Medium	2.34	.38	10	10.77	.01
	High	2.60	.51	13		
Setting	Low	1.96	.42	16		
	Medium	2.32	.56	10	9.34	.01
	High	2.66	.34	13		
Plot	Low	2.10	.54	16		
	Medium	2.66	.34	10	9.28	.01
	High	2.74	.34	13		
Communication	Low	2.14	.47	16		
	Medium	2.52	.34	10	9.20	.01
	High	2.77	.35	13		

BEST COPY AVAILABLE